DS203 PROJECT

Anousha Puvvala (23B2249)
Kashish Jain (23B0067)
Nandini Agrawal (23B0014)
Shivam Chaubey (23B1244)

Index

Problems Solved

1. Analyzed **115** MFCC files to organize them into **6** classes
2. Identified files containing **National Anthem**
3. Identified files containing solo songs by **Asha Bhosale**, **Kishore Kumar** and **Michael Jackson**

Executive Overview

Train data of 800 songs converted to 20 MFCC coefficients

Generated co-occurrence matrix for each coefficient

Calculated energy, entropy, homogeneity, contrast, correlation and average of coefficient

Used linear correlation heatmap and VIF for feature reduction

Performed PCA for 2D and 3D vizualisation

Created generalized classification models and best combination of train and test F1 score was chosen

Given Data

- **115** CSV files, one for each song
- **6** broad classes for the songs
- **20 rows** of MFCC coefficients
- **Variable number of columns** depending on song duration
- Sampling rate – 44100 Hz
- Hop size - 512

Mel Frequency Cepstral Coefficients (MFCC)

1. Take the Fourier Transform of (a windowed excerpt of) a signal
2. Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows or cosine overlapping windows
3. Take the logs of the powers at each of the mel frequencies
4. Take the discrete cosine transform of the list of mel log powers
5. The MFCCs are the amplitudes of the resulting spectrum

Description of MFCC

MFCC coefficients

First coefficient (MFCC 0)	Reflects **signal energy**; often omitted for volume invariance
Low-Order Coefficients (MFCC 1-5)	Captures **broad spectral shape**, essential for speech and tonal features in music.
Mid-Order Coefficients (MFCC 6-13)	Adds **finer frequency details**, useful for distinguishing speech sounds and musical timbres.
Higher-Order Coefficients (MFCC 14-20)	Generally less relevant, as it mainly **captures noise** and fine, often unnecessary details.

Our approach

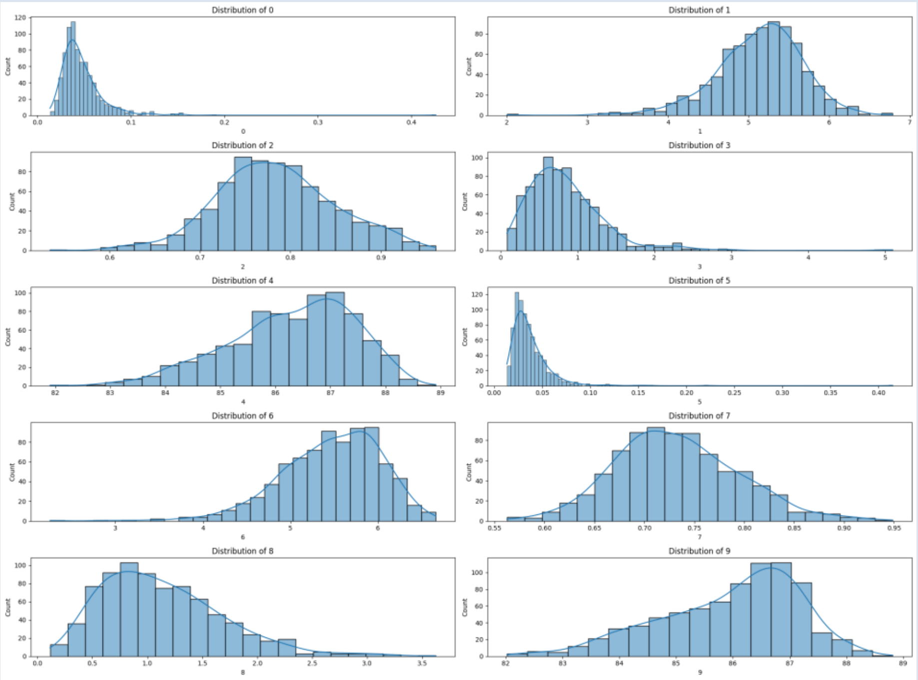
Dataset Preparation

Training data + Testing data:

Category	No. of Songs
Asha Bhosle	119
Michael Jackson	148
Lavani (Predominantly female)	83
Kishore Kumar	174
Bhav Geet (Both male and female)	150
National Anthem (Both male and female)	116

Further used librosa for generating the MFCC coefficient
Link to data - https://drive.google.com/drive/folders/19KQSR8-scUJL5xboj6XiK7EOg0AxnNUX?usp=drive_link

Exploratory Data Analysis



Summary Statistics:

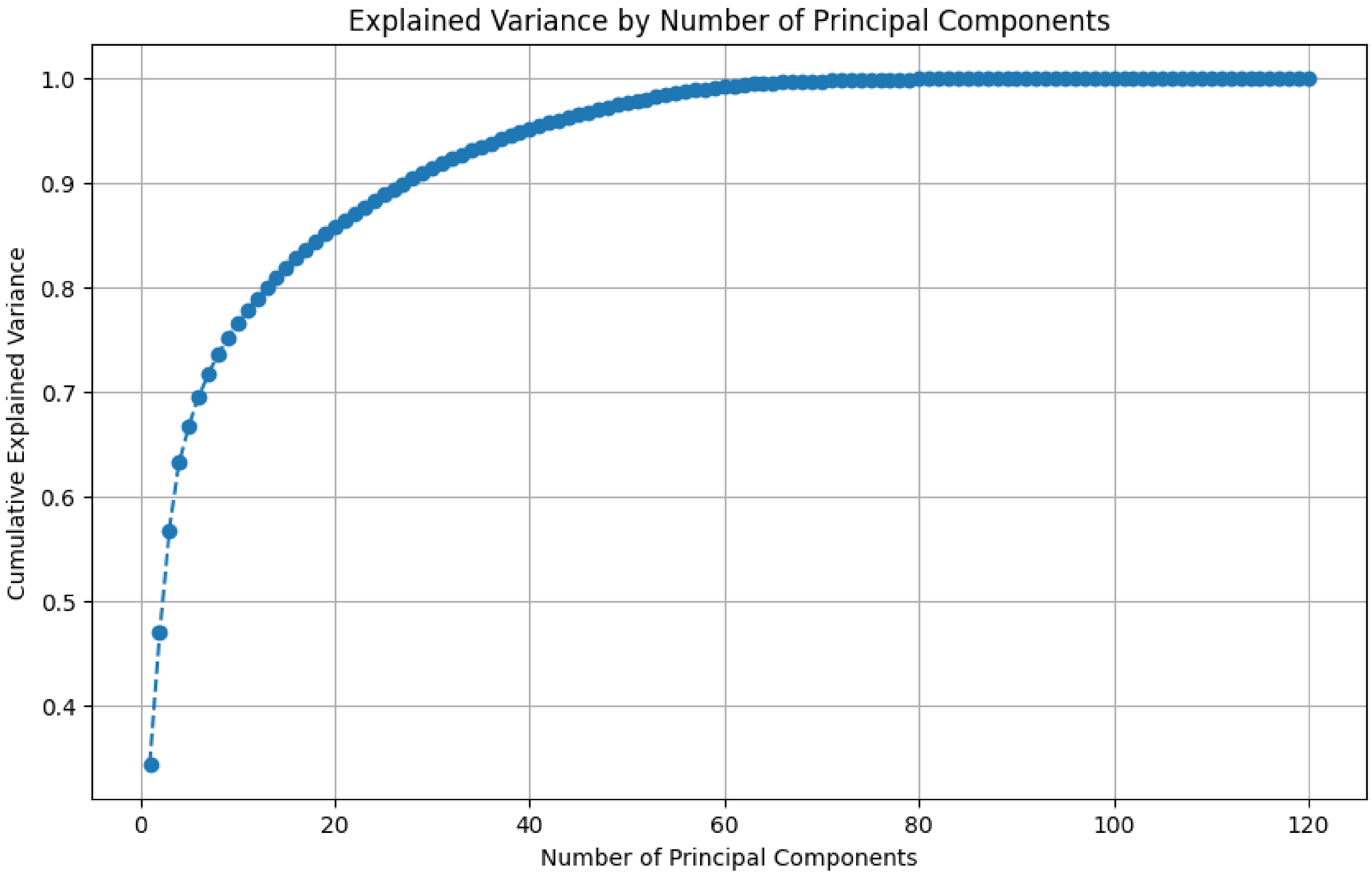
	0	1	2	3	4	5
count	790.000000	790.000000	790.000000	790.000000	790.000000	790.000000
mean	0.047297	5.130135	0.784519	0.847278	86.290650	0.037318
std	0.024805	0.537872	0.065979	0.514276	1.137576	0.023335
min	0.013646	2.002552	0.534124	0.083857	81.907032	0.013276
25%	0.033681	4.816255	0.740231	0.505170	85.601071	0.024849
50%	0.041663	5.186837	0.781925	0.756205	86.449434	0.032022
75%	0.054401	5.485996	0.826926	1.067249	87.160107	0.043888
max	0.426104	6.787183	0.960848	5.096290	88.913501	0.414843

Data Distribution: “Histograms showing the distribution of each feature. Helps identify skewness, spread, and possible outliers.”

Summary Statistics: “Statistical metrics for each feature, including mean, standard deviation, and quartiles. Useful for understanding central tendency and variability.”

PCA Analysis

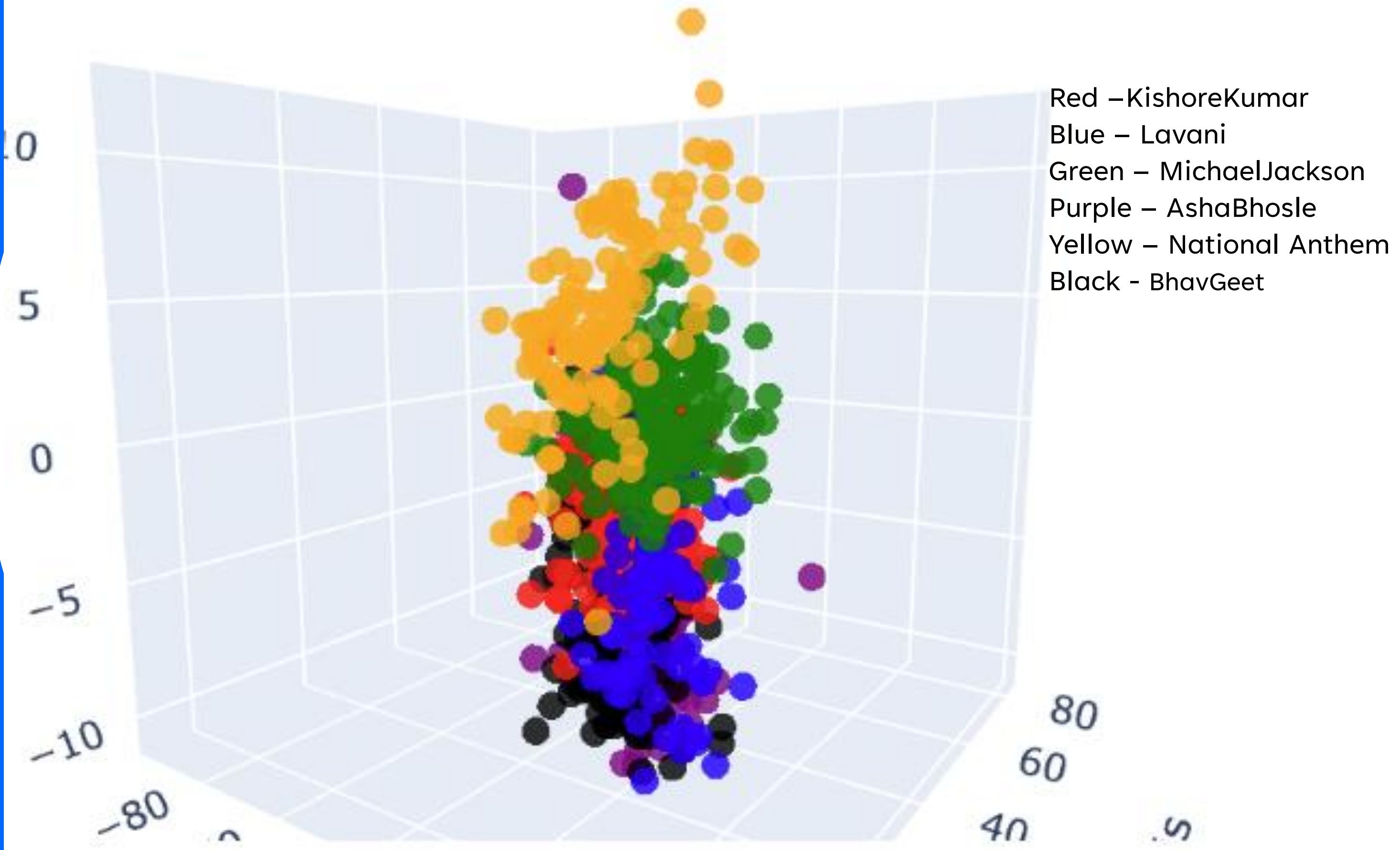
- Optimal number of components: 28 (to explain 90% variance)



PCA plot in 2D

- Cluster for National Anthem and Michael Jackson is relatively distinct compared to other 4
- Distinct clustering not observed for Asha Bhosle, Bhav Geet and Lavani due to similar nature of songs

PCA plot in 3D



- Clusters are more distinct in 3D especially for NationalAnthem and MichaelJackson
- Not optimum performance due to 28 components needed to explain 90% variance against the 3 used

Co-occurrence Matrix

- Each coefficient quantized considering $\mu i \pm k*\sigma i$ different quantization levels
  - μi - mean of coefficient
  - Σi - standard deviation of coefficient
- K varies from –2 to +2 with 0.25 step size
- 2 additional bins for values not fitting in any above mentioned bins
- Therefore total 18 bins
- 18 * 18 co-occurrence matrix M formed for each feature
- Mij represents the count of change of MFCC value from ith to jth bin in successive timeframes
- Normalized matrix P = M / total number of timeframes

Feature Engineering from Co-Occurrence Matrix

Extracted 5 features for every coefficient from its co-occurrence matrix -

- Energy
- Entropy
- Homogeneity
- Contrast
- Correlation

Energy

- Represents the uniformity or concentration of values in co-occurrence matrix
- Energy = $\sum_i \sum_j P(i,j)^2$
- Interpretation -
  - High energy might be associated with tracks with repetitive notes
  - Low energy indicates a uniform distribution consisting of varying sounds

Entropy

- Represents the randomness or disorder in the co-occurrence matrix
- Entropy = $-\sum_i \sum_j P(i,j) . \log(P(i,j))$
- Interpretation -
    - High entropy suggests complexity and variety in audio with no dominant patterns
    - Low entropy suggests predictability and stability with dominant and repetitive patterns

Homogeneity

- Represents how close the values are to the diagonal
- Homogeneity = $\sum_i \sum_j \frac{P(i,j)}{1+|i-j|}$
- Interpretation -
  - High homogeneity suggests smooth and consistent variation like in calm music
  - Low homogeneity suggests dynamic and varied values like in energetic music

Contrast

- Represents how far the values are from the diagonal
- Contrast = $\sum_i \sum_j P(i,j).(i - j)^2$
- Interpretation -
  - High contrast suggests dynamic and complex sounds with sharp transitions
  - Low contrast suggests stable, predictable sounds with less transitions

Correlation

- Represents linear relationship between patterns of bins or features across the timeframes.
- Correlation = $\frac{\left(\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y\right)}{\sigma_x \sigma_y}$
- Interpretation -
  - High positive or negative correlation suggests features in the frequency bins appear in similar or inverse pattern
  - Low or no correlation suggests independence of the features from each other

Summary of the Features -

- 5 features (energy, entropy, homogeneity, contrast, correlation) extracted for each MFCC
- Additional 20 features added representing the mean of MFCC to capture the general trend
- Size = 5 * 20 + 20 = 120 features

Model Selection

Classification models used

1. Logistic regression
2. Random forest classifier
3. Support Vector Machine (SVM) classifier
4. K-Nearest Neighbors (KNN) classifier

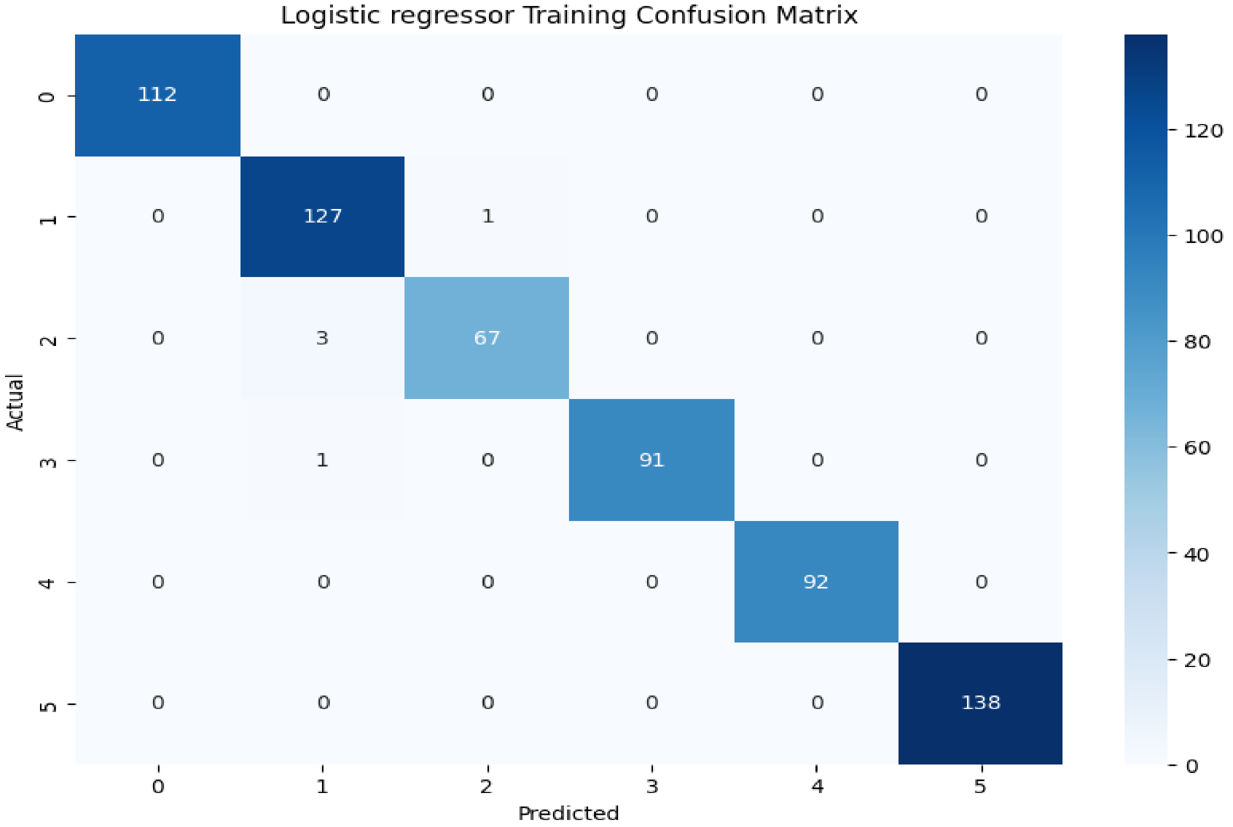
Metrics for 120 features

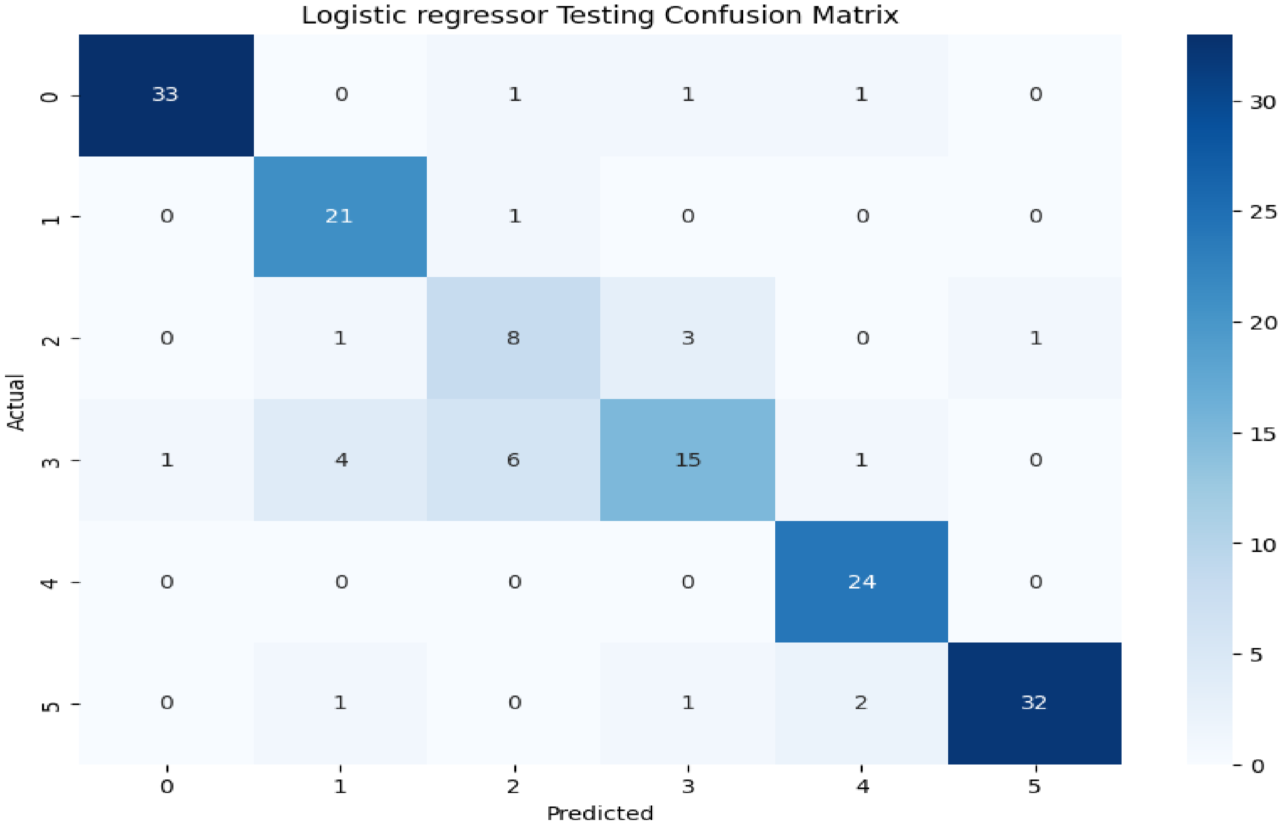
Model	Accuracy		Precision		Recall		F1 score	
	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	0.99	0.84	0.99	0.84	0.99	0.84	0.99	0.84
Random Forest Classifier	0.99	0.82	0.99	0.82	0.99	0.82	0.99	0.82
SVM Classifier	0.94	0.84	0.94	0.84	0.94	0.84	0.94	0.84
KNN Classifier	0.85	0.77	0.85	0.77	0.85	0.77	0.85	0.77

Metrics are good for logistic regression but there is difference between train and test f1 scores suggesting slight overfitting.

Logistic Regressor

labels=['MichaelJackson', 'BhavGeet', 'Lavani', 'AshaBhosle', 'NationalAnthem', 'KishoreKumar']





Model performs good on National Anthem, Michael Jackson, Kishore Kumar and Bhav Geet
Lavani and Asha Bhosle are not well classified as observed earlier

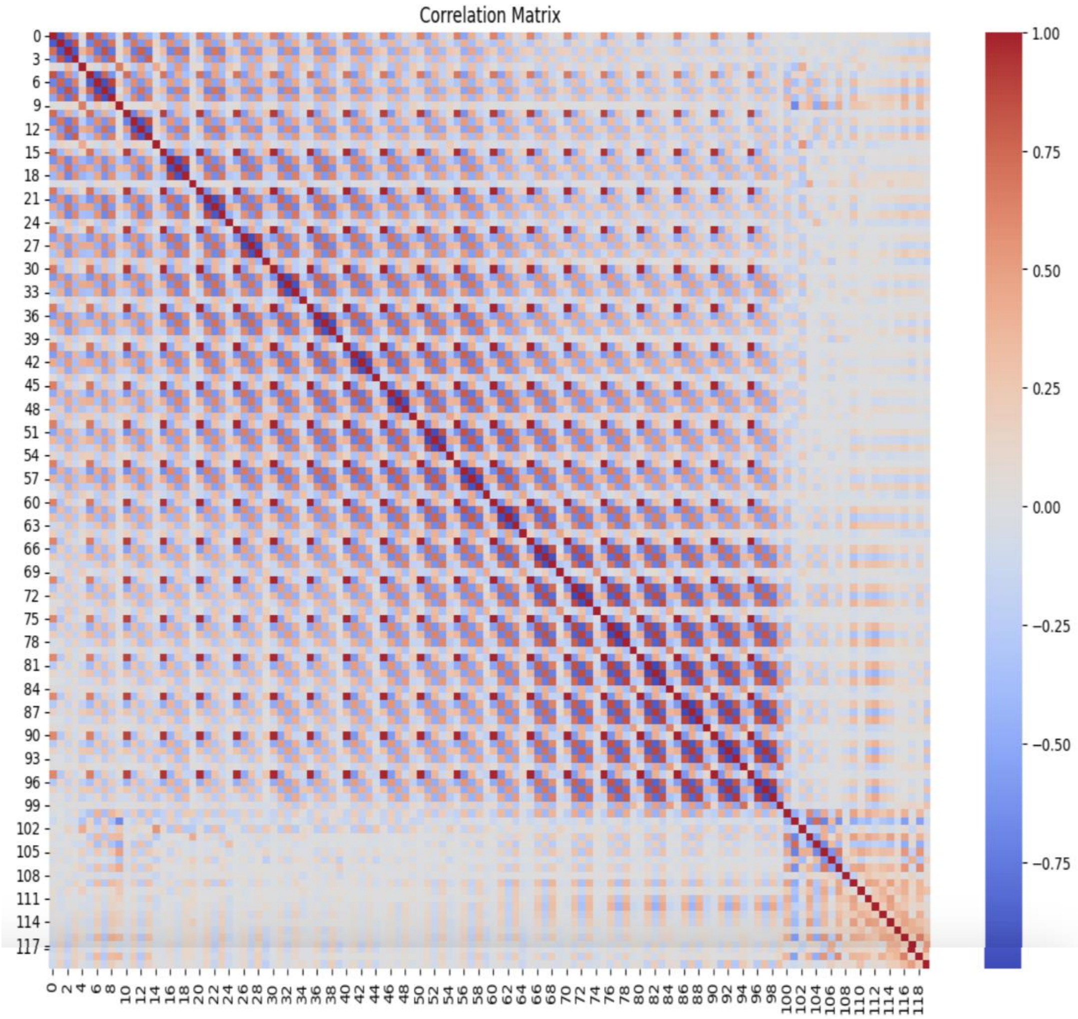
Feature Reduction

To reduce the overfit

Linear Correlation

Correlation Matrix for 120 Features

- High positive correlation observed between energy of MFCC after 3
- High negative correlation between contrast and homogeneity for all MFCC
- High negative correlation between entropy and homogeneity for all MFCC
- Columns with correlation greater than 0.9 were dropped



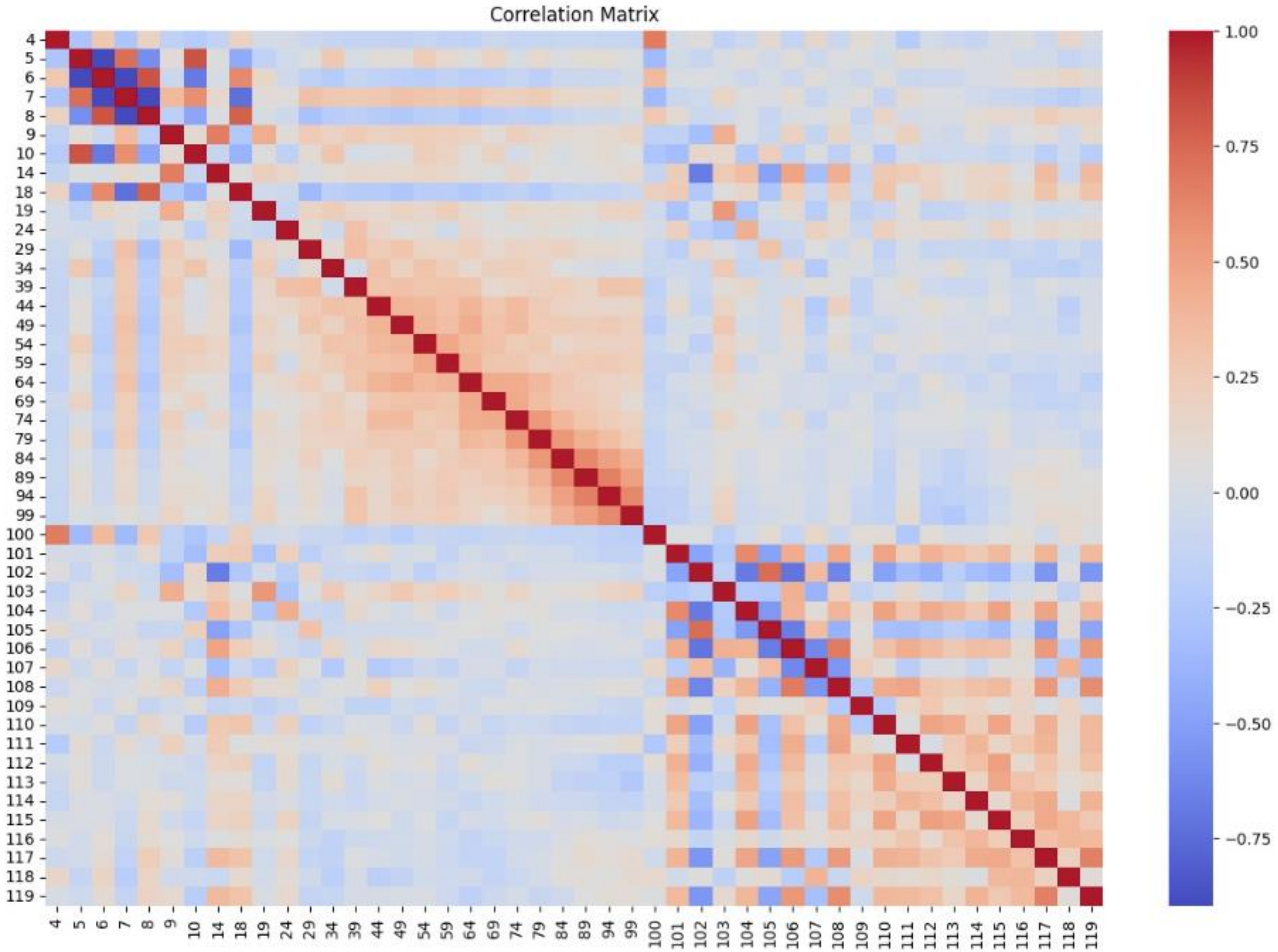
Metrics after Linear Correlation Feature Removal

Model	Accuracy		Precision		Recall		F1 score	
	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	0.91	0.76	0.91	0.76	0.99	0.82	0.99	0.82
Random Forest Classifier	1	0.77	1	0.77	1	0.77	1	0.77
SVM Classifier	0.94	0.81	0.94	0.81	0.94	0.81	0.94	0.81
KNN Classifier	0.79	0.74	0.79	0.74	0.79	0.74	0.79	0.74

Metrics have degraded even after choosing threshold as 0.9 suggesting that those features even after being highly correlated were important for prediction

Correlation Matrix for Reduced Dataframe

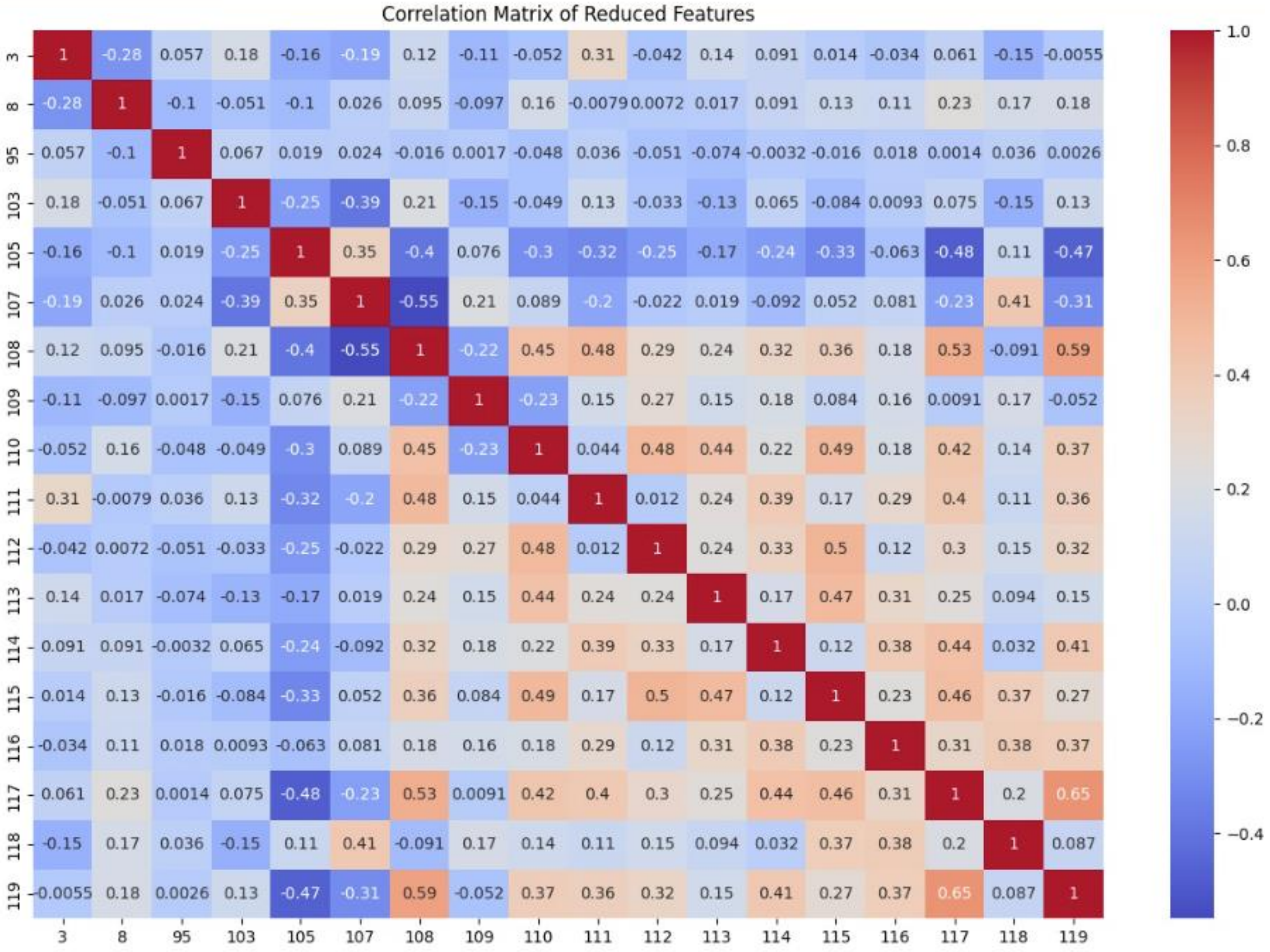
Most of the colours in the correlation matrix have lightened suggesting reduced correlation in the features



Variance Inflation Factor (VIF)

VIF

- Features with VIF > 5 were dropped
- Left with only 18 features after VIF
- Observed significant reduction in correlation after VIF



Metrics for VIF Features

Model	Accuracy		Precision		Recall		F1 score	
	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	0.82	0.73	0.82	0.73	0.82	0.73	0.82	0.73
Random Forest Classifier	1	0.75	1	0.75	1	0.75	1	0.75
SVM Classifier	0.89	0.75	0.89	0.75	0.89	0.75	0.89	0.75
KNN Classifier	0.83	0.70	0.83	0.70	0.83	0.70	0.83	0.70

This model is not performing better than the one with 120 features, so not using it
This might be due to some important features getting dropped due to the lack of domain knowledge

So our current model incorporates all the 120 features with logistic regression model!

Results – Asha Bhosle

	Names	label	prob
58	99-MFCC.csv	AshaBhosle	0.999644
82	72-MFCC.csv	AshaBhosle	0.999494
113	15-MFCC.csv	AshaBhosle	0.999235
77	79-MFCC.csv	AshaBhosle	0.998640
41	22-MFCC.csv	AshaBhosle	0.998441
87	04-MFCC.csv	AshaBhosle	0.997887
103	68-MFCC.csv	AshaBhosle	0.996754
51	115-MFCC.csv	AshaBhosle	0.996637
109	62-MFCC.csv	AshaBhosle	0.994592
9	06-MFCC.csv	AshaBhosle	0.992381
100	91-MFCC.csv	AshaBhosle	0.992233
74	106-MFCC.csv	AshaBhosle	0.991728
5	112-MFCC.csv	AshaBhosle	0.991131
7	94-MFCC.csv	AshaBhosle	0.986702
1	71-MFCC.csv	AshaBhosle	0.978359
93	40-MFCC.csv	AshaBhosle	0.975360
94	41-MFCC.csv	AshaBhosle	0.965171
10	33-MFCC.csv	AshaBhosle	0.939819
72	80-MFCC.csv	AshaBhosle	0.929546
107	56-MFCC.csv	AshaBhosle	0.915332
60	42-MFCC.csv	AshaBhosle	0.913580
34	85-MFCC.csv	AshaBhosle	0.911726
70	12-MFCC.csv	AshaBhosle	0.889369
73	107-MFCC.csv	AshaBhosle	0.857652
25	25-MFCC.csv	AshaBhosle	0.819362
39	102-MFCC.csv	AshaBhosle	0.810984
106	57-MFCC.csv	AshaBhosle	0.801313
50	48-MFCC.csv	AshaBhosle	0.784673
43	54-MFCC.csv	AshaBhosle	0.769737
90	36-MFCC.csv	AshaBhosle	0.762762
0	70-MFCC.csv	AshaBhosle	0.761068
42	109-MFCC.csv	AshaBhosle	0.716721
32	60-MFCC.csv	AshaBhosle	0.698282
18	82-MFCC.csv	AshaBhosle	0.693429
46	77-MFCC.csv	AshaBhosle	0.591587
64	51-MFCC.csv	AshaBhosle	0.515758
27	52-MFCC.csv	AshaBhosle	0.501192
61	27-MFCC.csv	AshaBhosle	0.488911
83	110-MFCC.csv	AshaBhosle	0.449466
76	31-MFCC.csv	AshaBhosle	0.421550

Results – Kishore Kumar

	Names	label	prob
91	09-MFCC.csv	KishoreKumar	0.999990
54	93-MFCC.csv	KishoreKumar	0.999602
35	84-MFCC.csv	KishoreKumar	0.999581
80	46-MFCC.csv	KishoreKumar	0.999553
68	65-MFCC.csv	KishoreKumar	0.998131
30	29-MFCC.csv	KishoreKumar	0.996665
66	18-MFCC.csv	KishoreKumar	0.993038
19	83-MFCC.csv	KishoreKumar	0.991757
63	50-MFCC.csv	KishoreKumar	0.981893
24	24-MFCC.csv	KishoreKumar	0.978370
23	58-MFCC.csv	KishoreKumar	0.977494
53	92-MFCC.csv	KishoreKumar	0.971297
11	32-MFCC.csv	KishoreKumar	0.960646
21	59-MFCC.csv	KishoreKumar	0.949963
99	02-MFCC.csv	KishoreKumar	0.932274
4	113-MFCC.csv	KishoreKumar	0.929502
14	67-MFCC.csv	KishoreKumar	0.920411
8	07-MFCC.csv	KishoreKumar	0.835260
65	19-MFCC.csv	KishoreKumar	0.817717
114	100-MFCC.csv	KishoreKumar	0.732054
89	37-MFCC.csv	KishoreKumar	0.714056
115	101-MFCC.csv	KishoreKumar	0.692197
45	55-MFCC.csv	KishoreKumar	0.665371
101	90-MFCC.csv	KishoreKumar	0.584161
108	63-MFCC.csv	KishoreKumar	0.580351
112	14-MFCC.csv	KishoreKumar	0.359290

Results – Michael Jackson

	Names	label	prob
49	114-MFCC.csv	MichaelJackson	0.999909
92	08-MFCC.csv	MichaelJackson	0.999234
78	78-MFCC.csv	MichaelJackson	0.998192
12	44-MFCC.csv	MichaelJackson	0.995792
13	45-MFCC.csv	MichaelJackson	0.990359
98	03-MFCC.csv	MichaelJackson	0.983023
38	103-MFCC.csv	MichaelJackson	0.980803
104	20-MFCC.csv	MichaelJackson	0.975018
55	34-MFCC.csv	MichaelJackson	0.972249
110	86-MFCC.csv	MichaelJackson	0.956539
95	74-MFCC.csv	MichaelJackson	0.925703
33	61-MFCC.csv	MichaelJackson	0.880585
26	53-MFCC.csv	MichaelJackson	0.592848
57	98-MFCC.csv	MichaelJackson	0.499570
56	35-MFCC.csv	MichaelJackson	0.471613
44	108-MFCC.csv	MichaelJackson	0.424682

Results – National Anthem

	Names	label	prob
37	16-MFCC.csv	NationalAnthem	0.999700
111	87-MFCC.csv	NationalAnthem	0.999574
96	75-MFCC.csv	NationalAnthem	0.998892
71	81-MFCC.csv	NationalAnthem	0.908467
6	95-MFCC.csv	NationalAnthem	0.904110
52	01-MFCC.csv	NationalAnthem	0.805935
36	17-MFCC.csv	NationalAnthem	0.560613

Results – National Anthem

	Names	label	prob
37	16-MFCC.csv	NationalAnthem	0.999700
111	87-MFCC.csv	NationalAnthem	0.999574
96	75-MFCC.csv	NationalAnthem	0.998892
71	81-MFCC.csv	NationalAnthem	0.908467
6	95-MFCC.csv	NationalAnthem	0.904110
52	01-MFCC.csv	NationalAnthem	0.805935
36	17-MFCC.csv	NationalAnthem	0.560613

Results – Lavani

	Names	label	prob
31	28-MFCC.csv	Lavani	0.996987
79	47-MFCC.csv	Lavani	0.991579
62	26-MFCC.csv	Lavani	0.990976
81	73-MFCC.csv	Lavani	0.946701
75	30-MFCC.csv	Lavani	0.928277
102	69-MFCC.csv	Lavani	0.902045
67	64-MFCC.csv	Lavani	0.896634
17	11-MFCC.csv	Lavani	0.887682
29	89-MFCC.csv	Lavani	0.883615
59	43-MFCC.csv	Lavani	0.746832
69	13-MFCC.csv	Lavani	0.615651
20	104-MFCC.csv	Lavani	0.601549
2	39-MFCC.csv	Lavani	0.572196
48	49-MFCC.csv	Lavani	0.542781

Results – Bhav Geet

	Names	label	prob
16	10-MFCC.csv	BhavGeet	0.999434
85	96-MFCC.csv	BhavGeet	0.985465
97	116-MFCC.csv	BhavGeet	0.985219
40	23-MFCC.csv	BhavGeet	0.972655
86	97-MFCC.csv	BhavGeet	0.838169
22	105-MFCC.csv	BhavGeet	0.798124
84	111-MFCC.csv	BhavGeet	0.778085
28	88-MFCC.csv	BhavGeet	0.666368
15	66-MFCC.csv	BhavGeet	0.628069
88	05-MFCC.csv	BhavGeet	0.557698
47	76-MFCC.csv	BhavGeet	0.529999
3	38-MFCC.csv	BhavGeet	0.442725
105	21-MFCC.csv	BhavGeet	0.395642

Results – Summary

Class	Top 5 files
Asha Bhosle	99, 72, 15, 79, 22
Kishore Kumar	9, 93, 84, 46, 65
Michael Jackson	114, 8, 78, 44, 45
National Anthem	16, 87, 75, 81, 95
Lavani	28, 47, 26, 73, 30
Bhav Geet	10, 96, 116, 23, 97

Challenges Faced

- Dataset creation being an exhaustive and time consuming process
- Not many sources to understand MFCC implementation clearly
- Understanding implementation and usage of co-occurrence matrix
- Finding the best model without it being highly overfit
- Implementing feature reduction to reduce overfit

Learnings

- Data collection and preparation
- Hands-on experience on a new domain
- Reviewing research papers to identify the best approach to our problem
- Effectively conveying the results and conclusions

Resources

- Research paper link (Majorly implemented this research paper) :

https://www.researchgate.net/publication/272620724_Music_Classification_based_on_MFCC_Variants_and_Amplitude_Variation_Pattern_A_Hierarchical_Approach

Evaluation Criteria - Results

- How many problems have been correctly solved?
  - 3
- Has there been any creative thinking and innovation while solving the problems?
  - Yes, the creation of a co-occurrence matrix was a completely new idea
- Quality of Feature Engineering / Feature Creation in terms of relevance to the problem
  - Features relevant as they express the energy, entropy, homogeneity, contrast, correlation for each feature which help in understanding the frequency variations in the songs

Evaluation Criteria - Process

- Are the solutions relevant, correctly applied, and backed up with proper metrics / reasons / explanations?
  - All the metrics for all the methods have been provided with reasons for their good or bad values
- Are the major steps of data analysis diligently followed and correctly applied and documented (wherever required / if applicable …)?
  - Done problem framing, data acquisition, data preparation, model planning, model building and metrics analysis which are the standard steps in data analysis. All the reasoning has been provided with all the major decisions taken.

Evaluation Criteria - Documentation

- Quality of presentation: Completeness and preciseness of the final slide deck; design and readability of the slides. Are all the above aspects covered in the presentation?
  - Tried to cover all the majority steps we had taken irrespective of whether they worked or not and backed it up with metrics and analysis. Tried writing short and concise points for effective communication

Thank you!!!